DOCUMENT RESUME

ED 214 938 TM 810 882

AUTHOR Mkandawire, Donton S.J.
TITLE The Utility of Predictive Assessment and the World of Work.
PUB DATE May 81
NOTE 18p.; Paper presented at the International Conference of the International Association for Educational Assessment (7th, Manila, Philippines, May 25-30, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Developing Nations; Evaluation Methods; *Measurement Techniques; Models; Occupational Tests; *Predictive Measurement; *Predictive Validity; *Selection; *Testing Problems; Testing Programs
IDENTIFIERS *Utility Theory

ABSTRACT

The problem of finding a meaningful way of demonstrating the benefit or satisfaction to be gained by using predictive assessment techniques to select or place individuals in work situations is explored. The urgent need of developing nations for appropriate occupational and educational selection procedures is presented and explained. A discussion on (1) the points to consider before implementing a selection program, (2) the concept of decision theory as applied to personnel selection, (3) the basic model often used in selection procedures, and (4) the precision model used for evaluationg selection tests is provided. In spite of the various procedures suggested to measure the worthiness of a test, it is concluded that it has not been possible to show, in a meaningful way, the practical gain to be achieved by investing in selection programs. (AL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED214938

THE UTILITY OF PREDICTIVE ASSESSMENT AND THE WORLD OF WORK

Donton S.J. Mkandawire, Ph.D.

Director, The Malawi Certificate Examination and Testing Board
P.O. Box 191, Zomba, Malawi

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D.S.J. Mkandawire

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 882

A paper presented at the Seventh International Conference of the International Association for Educational Assessment at the Philippine Convention Centre Manila, Philippines, May 25 - 30, 1981

THE UTILITY OF PREDICTIVE ASSESSMENT AND THE WORLD OF WORK

The art of assessment can be applied to both cognitive and affective domains. Cognitive measures attempt to assess maximum performance. It is an accepted phenomena that people differ, and assessing any of their characteristics with a measuring instrument results in a distribution of scores which is used in many different ways for decision-making purposes. It is also evident that the entire psychological testing movement got its impetus from an area of psychology concerned with individual differences. The existence of individual differences has provided the basis for assessment to maximize the objective of proper selection and placement in the world of work.

The word 'utility' is the satisfaction one gets from something, and 'utility value' would be a measure of that satisfaction. The art of assessment applied to the world of work will be primarily discussed in this paper in the context of those assessment procedures which are used to select or place individuals in work situations. The utility of assessment will therefore be the amount of satisfaction these procedures give clients or users of these assessment techniques.

* * * *

As the so-called third world countries have become independent, it is increasingly being realised that the acceleration of economic and technological advancement of these new, developing, and independent nations, has produced an urgent demand for appropriate occupational and educational selection procedures. Because of the limited resources available, the need for selection procedures has become more pressing. It has been realised that it is impossible to educate and train all individuals who might be eligible for and benefit

from such programmes. Beisheuvel (1962) and Taylor (1962) have described the acute shortage of trained manpower and resources. Therefore, it is important to obtain maximum utility from the limited educational opportunities and trained manpower presently available.

Practically all developing nations have accepted, in their economic plans, the principle of maximal manpower utilisation. This principle has become a directional force so that planning activities are based on the philosophy that individuals capable of work should be placed in jobs or training which demand full and efficient use of their talents and will consequently bring the greatest benefit to the society. Educators, psychometricians, and psychologists have come under great pressure from policy makers to introduce allocation methods which are consistent with this viewpoint. Clarke (1959), together with many others who have been concerned with this position, has suggested that, given the situation, the only way developing nations can meet their aspirations is to develop selection procedures which will correctly identify individuals for the relatively few places that are available in both educational institutions and industries. As a consequence, the fields of educational and psychological measurement will, for a long time to come, be preoccupied with the use of tests both in occupational and educational selection.

Within the various developing nations, governments are establishing testing programmes whose major objectives are to develop tests which can be used to select applicants for educational institutions and industries. They have adopted the selection-rejection model which follows the standard format: applicants are given some form of aptitude test and those who perform above a determined cut-off point are selected; the rest are rejected. Cronbach and Gleser (1965) call this a non-sequential testing strategy or a single stage testing strategy. This model assumes that those selected will be given similar

(fixed) treatment whether it be an educational programme or type of job they will be required to do.

In their efforts to establish selection procedures, the various developing nations have found it necessary to follow a validation model which requires predictive validity research studies on all test batteries before they are implemented. The worthiness of a selection test battery has been determined principally by the extent to which predictor tests included in the selection plan correlate with the criterion.

Professional workers responsible for developing selection tests have used expectancy charts and validity coefficients to express the utility of these assessment procedures. However, administrators in government have demanded to know what gains would be achieved by introducing selection programmes in relation to what it costs to establish them. Private industries have been reluctant to adopt the use of tests unless net gain in their profits can be demonstrated with the use of selection tests.

Although governments and other organisations want better selection procedures, which they hope will bring about the greatest benefit from their human resources, they do not have a meaningful way of evaluating selection tests.

* * * *

Selection Programmes

Hill (1971) has pointed out that an institution contemplating the establishment of a selection programme must make certain commitments:

(a) An institution must know what benefits it expects from the programme. Benefits may vary from institution to institution. In industry, selection is expected to contribute to the organisation's profit-making through the recommendation of people to be hired who will on the average bring most satisfaction to the industry. In an educational

institution the immediate benefit from the selection programme might be the admittance of applicants who have a high probability of graduating. However, the ultimate benefit in most cases is to have people who will be productive at work after they graduate.

(b) The institution must decide whether the treatment, once applicants have been selected, will be *fixed* or will be *adaptive*. In a fixed treatment everybody does similar types of work or goes through the same educational programme. An adaptive treatment is one in which the type of job or educational programme is modified to accommodate the characteristics of the selected applicants in order to achieve the desired outcomes. In most developing countries, due to limited financial resources, the people who are selected receive the same treatment. The selection-rejection model used is of the fixed treatment type.

(c) The nature of the applicant population must be determined. It will be necessary to know in advance whether everybody applying can be considered, or whether restrictions will have to be made on the applicant population.

(d) The programme must decide what selection instruments will be used in making the selection decisions. In most instances, educational and psychological tests have been used.

(e) It will also be necessary to know how the measures on the instruments will be combined. The regression model has been widely employed to combine measures.

(f) Finally, a very important decision must be made regarding the cutoff point on the measures necessary to accrue the expected payoff from those who are selected.

* * * *

A decision problem has been described as a situation in which an individual is confronted with alternative courses of action from which he must choose. Wald (1950) and Adams (1960) have described decision situations as having certain elements. First, there will generally be a set of alternative actions available from which the decision-maker must choose. Second, there will be a set of possible states of nature which are beyond the decision-maker's control. For each decision taken there will be a set of possible outcomes or consequences which are of interest to the decision-maker. The decision-maker must evaluate each outcome and assign to it a utility value (measure of satisfaction) in order of preference. Which of the outcomes occurs in a given instance will largely depend on both the action taken and the state of nature. Since decision-makers will prefer some outcomes to others, it is necessary to come up with assessment procedures which will show practical benefits to be derived from actions producing different outcomes.

Cronbach and Gleser (1965) have emphasised the fact that personnel decisions represent a special case within the statistical decision theory as originally developed by Wald (1950). This can best be demonstrated in the selection of applicants for jobs (Cronbach, 1971). Decisions are made whenever there is a need to assign one or more applicants to treatments, and the eventual outcome of such assignment is of major importance to the institution concerned. The term 'treatment' takes on a very wide meaning which includes such things as training programmes, counselling procedures, or occupational assignments (Cronbach and Gleser, 1965). As they see it, a decision in personnel work concerns what treatment to give one or more persons: Should a person be accepted for training or hired for a job? For what job should he be trained? To what counselling, therapy, or educational programme should he be assigned? Every personnel decision involves assigning

individuals to an appropriate treatment.

Cronbach and Gleser (1965) see three types of personnel decisions frequently encountered. First is the selection decision problem in which an institution has to decide to accept some applicants and assign them to treatment, and to reject others. Examples of this would be hiring or rejecting applicants for employment, or admitting some applicants to an academic institution and rejecting others who do not qualify.

Second is the classification decision problem in which there are two or more paths along which assignment could be made. In classification there is no rejection as an acceptable alternative. Rather, a decision is made regarding to which one of several available treatments applicants should be assigned. Rulon, Tiedeman, Tatsuoka, and Langmuir (1967) have studied and described the classification problem in depth. The standard classification problem entails assigning each applicant to one of several sub-groups. In industry this is often by assigning applicants to different job assignments, etc. A classical example of classification is in the military service where decision-makers are typically concerned with how to assign recruits once they have completed basic training (Cronbach and Gleser, 1965). Decisions have to be made as to how each recruit must be assigned to one of a number of service schools or to one of a number of on-the-job training programmes.

The third category stated by Cronbach and Gleser (1965) is the placement problem. Here, on the basis of a selection instrument, decision-makers assign applicants to different levels of the same treatment. A classical example of placement would be the grouping of students into several classes in accordance with their demonstrated ability on some intelligence test or other aptitude measure. Decision-makers have used tests as instruments to aid them in making rational decisions. Statistical regression analysis has been the basic model for prediction for most of the types of decisions I have so far described.

Cronbach and Gleser (1965) have distinguished between institutional decisions and individual decisions.

> *Institutional decisions such as choice of employees may be distinguished from individual decisions such as choice of a vocation. In the typical 'institutional' decision, a single person makes a large number of comparable decisions. The 'individual' decision is one in which the choice confronting the decision maker will rarely or never recur.*

Since, in the case of selection, institutional decisions are concerned with either accepting or rejecting an applicant, a decision-maker will want to make those decisions which will accrue as much benefit as possible to the institution he represents. Unfortunately, not all of the necessary information is available to the decision-maker. One can only hope to adopt a strategy which will work best on the average.

Wiggins (1973) stated that:

> *Given a choice between acts (placements) in the face of uncertain outcomes (performance), the effective decision maker behaves in such a fashion as to maximize the values (profits) of the institution he represents. In the role of the employee, a decision maker may accept or be indifferent to the values implicit in institutional gain, but if he is an applied scientist, he is thoroughly committed to the principle of maximization. Assessment psychologists who drag their feet or are otherwise indifferent to the principle of maximization are implicitly supporting irrational behaviour in areas of great social importance.*

Psychological and educational tests, which are developed for selection purposes, are supposed to aid the decision-maker in coming up with institutional decisions which will be of most benefit to the institution. Cronbach and Gleser (1965) deplore the fact that the classical test theory places too much emphasis on the accuracy of measurement, and not enough emphasis on the theory of test construction and usage which can then best serve the decision-maker in making rational decisions. While accepting the fact that accuracy is an essential aspect of measurement, they also disagree with Hull (1928)

who asserted that *'the ultimate purpose of using aptitude tests is to estimate or forecast aptitudes from scores'*. Instead they think that *'the ultimate purpose of any personnel testing is to arrive at qualitative decisions'*.

Wiggins (1973) also remarked that the emphasis on accurate prediction in measurement does not take into consideration the fact that decisions based on the same measurements might have different value outcomes. The evaluated outcomes of some selection decisions are more important than others.

* * * *

Basic Selection Model

The basic model which is often used in selection procedures is diagrammed in Figure 1.

The establishment of a criterion is a basic issue in the prediction of any human behaviour. The criterion has been defined in different ways. One definition is that by Guion (1965) who defines it as *'that which is to be predicted'*. That which is to be predicted in selection-rejection problems can be either good grades or good job performances. In essence, it is a measure of the degree of accomplishment in one form or another. When the criterion measure is obtained close to the time at which prediction measures are obtained, it is referred to as a *proximal criterion*. If the criterion measure is obtained at some later date, it is referred to as a *distal criterion*. When the criterion measure is obtained at the same time the prediction measures are determined, then it is referred to as an *immediate criterion* (Thorndike, 1949). The distal (ultimate) criterion, although the most difficult to determine, is often what is of real interest to society.

Once the criterion measure has been established it is necessary to select measuring instruments which will predict the criterion.

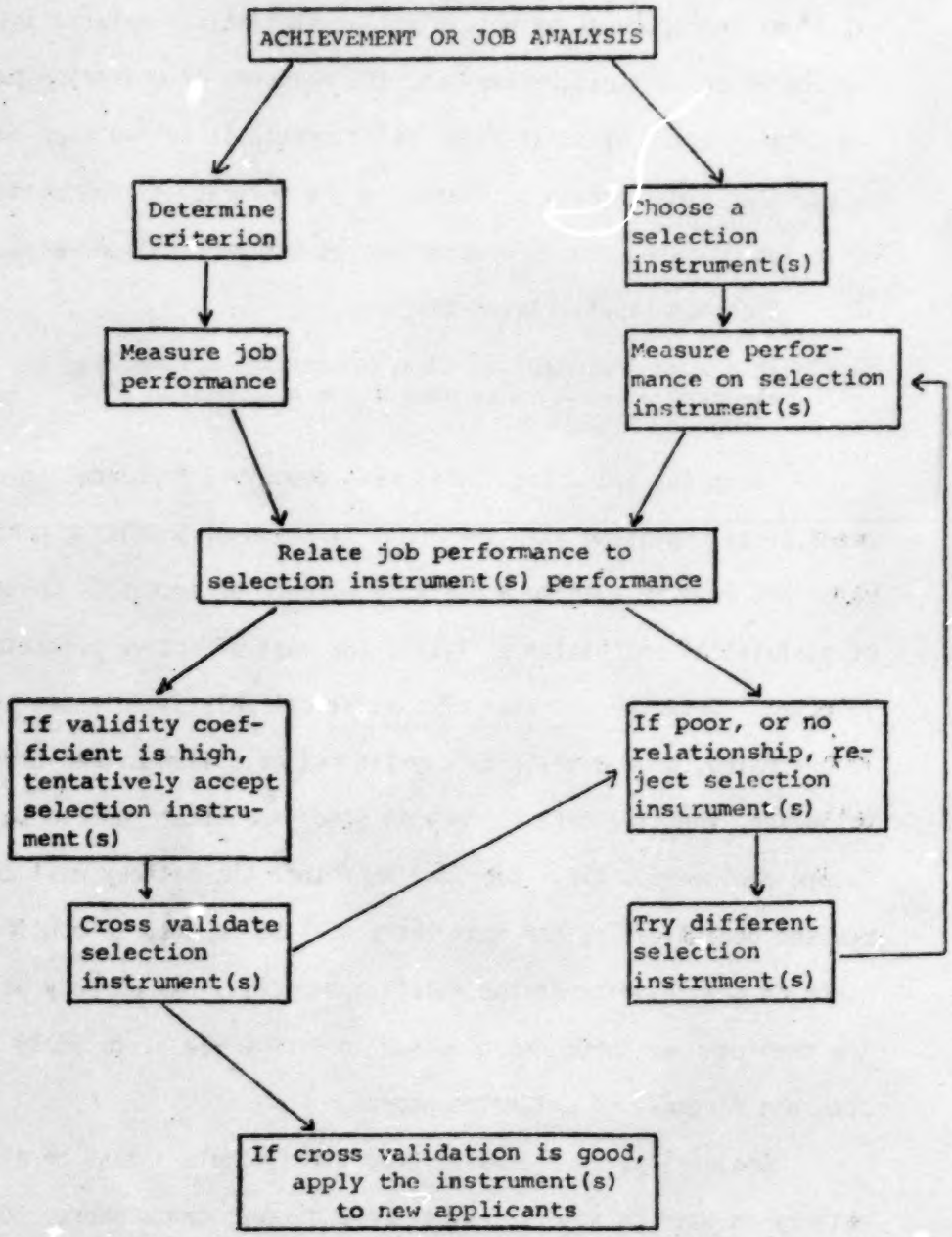
ACHIEVEMENT OR JOB ANALYSIS
Determine criterion
Choose a selection instrument(s)
Measure job performance
Measure performance on selection instrument(s)
Relate job performance to selection instrument(s) performance
If validity coefficient is high, tentatively accept selection instrument(s)
If poor, or no relationship, reject selection instrument(s)
Cross validate selection instrument(s)
Try different selection instrument(s)
If cross validation is good, apply the instrument(s) to new applicants

FIGURE 1: SCHEMA OF BASIC SELECTION MODEL

Adapted from Blum and Naylor, INDUSTRIAL PSYCHOLOGY: ITS THEORETICAL AND SOCIAL FOUNDATIONS, 1968,

While the initial identification of potentially useful tests is made on rational grounds, the validation of the selection test battery is empirical and proceeds by obtaining the statistical relationship between the scores on the test battery and the measures of criterion performance. The determination of statistical relationship is called *test battery validation.* The process of examining the accuracy of the prediction of criterion measures from predictor scores has been called validation.

Cronbach (1971) stated that:

> *In the narrow view, validation consists of checking the test score against some other observation that serves as criterion.*

Thorndike and Hagen (1969) have described personnel selection research as beginning with the identification of promising predictors which are then related to a suitable measure of success. On the basis of statistical regression analysis, the most effective predictors are selected. Once the evidence of statistical relationship has been demonstrated, a good validity coefficient established, and cross validated, then the test battery is used year after year to select future employees. (This does not mean that the battery will not be revised periodically, nor that there will be no need to continually check on the validity of the existing battery.) As already pointed out, the predictor measures are combined in such a way as to yield the most accurate forecast of criterion scores.

Ordinarily, in the basic prediction model, a test or a test battery is used to select from a group of applicants those individuals who are likely to perform efficiently on the job or in school. This is the reason why some measure of job success or school achievement is used as criterion. The worthiness or value of a test in this context has been judged traditionally simply by the accuracy of prediction of this criterion. The accuracy of prediction assumes that the predictor

measures are used as a substitute for the criterion measures because the criterion itself would presumably be used as a selector if expense, time, feasibility, and effort were not prohibitive.

Brogden (1946) stated that:

> *A measure of efficiency of selection by the predictor should indicate directly the proportion of maximum saving that is actually obtained with the use of the given test where the maximum saving is that obtained by selecting on the criterion.*

In employment and educational selection, tests are used as selection instruments. Tests have been recommended to users on the basis that measurements obtained from tests provide relatively precise estimates of an applicant's criterion performance. After all, so the argument goes, the objective of prediction, whenever selection is done, is to minimize as much as possible the predictive error in forecasting of criterion performance.

Richardson (1944), trying to clarify the interpretation of validity coefficient, noted that when the value of a test is based on the correlation coefficient between test scores and the criterion variable, the validity coefficient does not lend itself to an interpretation which is both scientifically sound and meaningful to the practical test user.

* * * *

Precision Model for Evaluating Selection Tests

The precision model evaluates tests by examining how accurate applicants' selection test scores forecast their criterion performance. The degree of precision in predicting criterion scores has been used as a measure of how useful selection tests are.

Cronbach and Gleser (1965), who, incidentally, were among the first to introduce the concept of decision theory to personnel selection,

noted that reliability and the accurate prediction of criterion measure is just one of the components in the decision-making process. The most important consideration is the outcome of decisions made as a result of the selection process.

Wiggins (1973) stated that:

> *Measurement and prediction are simply technical components of a system designed to make decisions about the assignment of personnel to jobs or treatments. Although measurement and prediction may be evaluated by formal test criteria such as reliability and validity, the outcomes of personnel decisions must be evaluated in terms of their consequence on individuals and institutions within the society.*

It has been suggested by Biesheuvel (1952) that selection should not be viewed from a prediction viewpoint only, but also in the wider context of more fully developed methods of recruitment, job requirements, work organisation, and factors which determine motivation within them. If all these variables could be considered, use of tests in selection, according to Biesheuvel, could be viewed as one of the many elements in the complex of manpower utilisation. Because of these variables, one should not appraise the worth of tests without taking into account the elements that determine the functioning of the system. Schwarz (1961) sees the value of predictor tests as being the improvements in the accuracy of decisions that would result.

Brogden (1946) demonstrated mathematically that in any selection situation the maximum gain in selection would occur when the criterion itself was used to select people. He therefore suggested that one measure of the worthiness of a test would be how much saving would be obtained with the use of a test in relation to the maximum gain. He called this the *index of selection efficiency*. A zero value would thus be obtained by random selection.

When selection testing programmes are established, test developers

are interested in showing that the applicants selected using the selection tests which have been developed, will result in some saving over random selection, or whatever procedure was used before tests were introduced. The burden to convince prospective test users of the benefits of tests has not been an easy one. One can only justify the use of a selection test if some benefit can be demonstrated. Since demonstration of the benefit obtained from the introduction of selection tests is of practical interest, psychometricians have suggested indices of evaluating tests.

The indices which have been suggested have relied heavily on statistical bases for describing the value of tests. The most common way of presenting evidence concerning the usefulness of the test, has been to report its validity coefficient. When the validity coefficient is high and statistically significant it has been assumed that the test can be useful as a means of evaluating applicants. When the validity is low and statistically insignificant, the test has been said to be ineffective. Taylor and Russell (1939) were among the first to point out that this criterion for test evaluation is misleading because gains resulting from the use of tests are also functions of other variables. Attention was then turned to the interpretation of the validity coefficient itself in determining the utility of the assessment procedures. Indices of forecasting efficiency were introduced which compares the standard error of scores when selection was by random process to the standard error when selection tests with validity coefficients were used. This led to the introduction of coefficient of alienation which emphasised that only tests with high validities had any practical value. Coefficient of determination (r^2ye) has also been used as a measure of the utility of tests. Inspite of all these procedures it has not been possible, in my view, to show the practical

gain, in a meaningful way, to employers when they are concerned with profits which are likely to accrue as a result of the investment they make in testing programmes. As a matter of fact, Cronbach and Gleser (1965) demonstrated mathematically that *'neither the index of forecasting efficiency nor coefficient of determination is directly related to utility when payoff functions are linear'*.

* * * *

Since this conference is supposed to focus on the techniques rather than general theories, maybe through our deliberations and/or experience of others in this room, or future efforts of IAEA, we can come up with predictive assessment procedures which can be used to show practical benefits to clients in the defence of the utility of assessment and the world of work. Thank you.

* * * *

REFERENCES

Adams, E.W., A survey of Bernaullian utilities and applications, in H. Solomon (Ed.), *Mathematical thinking in the measurement of behaviour*. Illinois: Free Press, 1960, pp. 151-268.

Biesheuvel, S. The study of African ability: the intellectual potentialities of Africans. *African Studies*, 1952, 2, 45-57.

Biesheuvel, S. The detection and fostering of ability among under-developed peoples. *Yearbook of Education*. London: Evans Bros. Ltd., 1962, 337-52, 541.

Brogden, H.E. On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology*, 1946, 37, 65-76.

Clarke, M.G. Secondary school selection in Northern Rhodesia. *Overseas Education*, 1959, 31, 99-101.

Cronbach, L.J. Test validation, in R.L. Thorndike (Ed.), *Educational Measurement (2nd Edition)*. Washington, D.C.: American Council on Education, 1971, pp. 443-507.

Cronbach, L.J. & Gleser G.C. *Psychological tests and personnel decisions (2nd Edition)*. Urbana: University of Illinois Press, 1965.

Guion, R.M. *Personnel testing*. New York: McGraw-Hill Book Co. Inc., 1965.

Hill, J.R. Use of measurement in selection and placement, in R.L. Thorndike (Ed.), *Educational measurement (2nd Edition)*. Washington, D.C.: American Council on Education, 1971, pp. 680-732.

Hull, C.L. *Aptitude testing*. Yonkers: World Book, 1928.

Richardson, M.W. The interpretation of a test validity coefficient in terms of increased efficiency of a selected group of personnel. *Psychometrika*, 1944, 9, 245-248.

Rulon, P.J., Tiedeman, D.V., Tatsuoka, M. & Langmuir, C. *Multivariate statistics for personnel classification*. New York: John Wiley & Sons, 1967.

Schwarz, P. *Aptitude tests for use in developing nations*. Pittsburgh: American Institutes for Research, 1961.

Taylor, A. *Educational and occupational selection in West Africa*. London: Oxford University Press, 1962.

Taylor, H.C. & Russell, J.T. The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology*, 1939, 23, 565-578.

Thorndike, R.L. *Personnel selection: tests and measurement techniques*. New York: John Wiley & Sons, 1949.

Thorndike, R.L. & Hagen, E. *Measurement and evaluation in psychology and education (3rd Edition).* New York: John Wiley & Sons, 1969.

Wald, A. *Statistical decision functions.* New York: John Wiley & Sons, 1950.

Wiggins, J.S. *Personality and prediction: principles of personality assessment.* Massachusetts: Addison-Wesley Publishing Co. Inc., 1973.